Temporality as Seen through Translation: A Case Study on Hindi Texts

Temporality as Seen through Translation: A Case Study on Hindi Texts

Sabyasachi Kamila† sabysachi.pcs16@iitp.ac.in
Sukanta Sen† sukanta.pcs15@iitp.ac.in
Mohammad Hasanuzzaman* hasanuzzaman.im@gmail.com
Asif Ekbal† asif@iitp.ac.in
Andy Way* andy.way@adaptcentre.ie
Pushpak Bhattacharyya† pb@iitp.ac.in
†Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, India
*ADAPT Centre, School of Computing, Dublin City University, Dublin, Ireland

Abstract

Temporality has significantly contributed to various aspects of Natural Language Processing applications. In this paper, we determine the extent to which temporal orientation is preserved when a sentence is translated manually and automatically from the Hindi language to the English language. We show that the manually and automatically identified temporal orientation in English translated (both manual and automatic) sentences provides a good match with the temporal orientation of the Hindi texts. We also find that the task of manual temporal annotation becomes difficult in the translated texts while the automatic temporal processing system manages to correctly capture temporal information from the translations.

1 Introduction

There is a considerable academic and commercial interest in processing time information in text, where that information is expressed either explicitly, implicitly, or connotatively. Recognizing such information and exploiting it for Natural Language Processing (NLP) and Information Retrieval (IR) tasks are important features that can significantly improve the functionality of NLP/IR applications such as event timeline generation, question answering, and automatic summarization (Mani et al., 2005; Campos et al., 2014).

Earlier studies on temporal information processing have mainly focused on identifying temporal expressions fostered by TempEval challenges (Verhagen et al., 2010; UzZaman et al., 2013). More recently, new trends have emerged in the context of human temporal orientation, which refers to individual differences in the relative emphasis one places on the past, present, or future (Zimbardo and Boyd, 2015). Past studies have established consistent links between temporal orientation and demographic factors such as age, sex, gender, education, and psychological traits (Webley and Nyhus, 2006; Adams and Nettle, 2009; Schwartz et al., 2013; Zimbardo and Boyd, 2015). In order to create a measure of user-level human temporal orientation measure, a message-level[1] temporal

[1]Only the English message is considered from microblogs.

classifier of past, present, and future is used. For instance, the following microblog post "*can't wait to get a pint tonight*" is automatically tagged as *future* by the temporal classifier. Successful features include timexes, specific temporal (past, present, future) words from a commercial dictionary, but also *n*-grams.

Many tasks in NLP are language-dependent, i.e. the same approach cannot be applied across different languages. In this case, one naive way of temporality detection is to translate the text automatically into the desired language and then apply any temporality detector system. However, Machine Translation (MT) itself is a challenging task and often the meaning, sentiment (Salameh et al., 2015; Lohar et al., 2017), temporarily of a text may not be preserved in the target language.

In this paper, we discuss the degree of preservation of underlying temporal orientation of a sentence when it is translated from Hindi to English. We use Hindi and English temporality analysis systems (described in Section 6.2) as well as a state-of-the-art Hindi-to-English translation system (Koehn et al., 2003). From our experiments, we attempt to analyze all the possible cases and answer the following questions:

1. What is the accuracy of temporality prediction by an English temporality analysis system when *Hindi* texts are translated into *English*?

2. How good are these predictions when compared to the Hindi temporality system?

3. What is the loss in the temporality predictability when translating the Hindi text into English automatically vs. manually?

4. What is the difficulty level to determine temporality by humans in automatically translated texts from Hindi to English?

5. Which is better in detecting temporality of the Hindi text in the translated English text: (a) human temporal annotation of the translated text or (b) automatic temporality analysis of the translated text?

We know that linguistic divergences between a pair of languages play significant role while translating from one language to the other language, and hence it has a significant impact on the accuracy of an automatic computational model. Our specific goal here is to analyse the temporality predictability of the Hindi text after translation. However, we confer that similar experiments can be validated for other language pairs to determine the impact of translation on temporality.

We show the percentage of temporality preservation in the translated English sentences, with respect to the temporality of Hindi sentences. We also show that both manual and automatic translations produce a change of temporality from that of the Hindi texts; *past* and *present* sentences tends to be translated into sentences of *future* time. Our further analysis shows that some characteristics in the automatically translated text mislead humans to correctly detect the temporality of the source text, and some of those were correctly classified by the automatic temporal analysis system.

Our contributions can be summarized as follows: i). to the best of our knowledge this is the first systematic attempt which presents a study whether temporality is preserved after translation; ii). we prepare a benchmark setup by creating three annotated datasets- Hindi texts, manual and automatic translated English texts labeled with three temporal classes, namely *past*, *present* and *future*; and iii). detecting the change of temporality in both manually a automatically translated sentences.

2 Related Works

Temporality has recently received increased attention in NLP and IR. The introduction of the TempEval task (Verhagen et al., 2009) and subsequent challenges (TempEval-2 and -3) in the Semantic Evaluation workshop series have clearly established the importance of time in dealing with different NLP tasks.

According to Metzger (2007), time is one of the key five aspects that determines a document credibility besides relevance, accuracy, objectivity and coverage. Given this, the value of information or its quality is intrinsically time-dependent. As a consequence, a new research field called Temporal Information Retrieval (T-IR) has emerged and deals with all classical IR tasks such as crawling (Kulkarni et al., 2011), indexing (Anand et al., 2012) or ranking (Kanhabua et al., 2011) from the viewpoint of time. From an application perspective of T-IR, Campos et al. (2014) proposed a solution for temporal classification of queries by identifying the top relevant dates in web snippets with respect to a given implicit temporal query, with temporal disambiguation performed through a distributional metric called GTE. Competitions like the NTCIR-11 Temporalia task (Joho et al., 2014) further pushed this idea and proposed to distinguish whether a given query is related to *past*, *recency*, *future* or *atemporal*. In order to push forward further research in temporal NLP and IR, Dias et al. (2014) developed TempoWordNet (TWn), an extension of WordNet (Miller, 1995), where each synset is augmented with its temporal connotation (past, present, future, or atemporal). Same kind of approach was followed for Hindi to create a lexical resource, namely TempoHindiWordNet (Pawar et al., 2016).

At the same time, there has been quite a few works on MT involving the Hindi-English language pair. Most of these systems aim to translate from English to Hindi or Indian languages (Dave et al., 2001; Sinha and Jain, 2003; Sinha and Thakur, 2005; Ananthakrishnan et al., 2006; Dungarwal et al., 2014; Sachdeva et al., 2014; Sen et al., 2016). One of the major challenges in MT between Hindi to English is the syntactic divergence. English follows the word order of Subject-Verb-Object (SVO) whereas Hindi follows Subject-Object-Verb (SOV). Ramanathan et al. (2008) have shown that simple syntactic transformation of the English language to meet the syntax of Hindi can improve translation quality. For our Hindi-English translation system, we follow the standard phrase based statistical MT (Koehn et al., 2003) approach.

3 Methodology Overview

We present our experimental setup to study the impact of translation on temporality, as follows:

1. Collect a Hindi dataset (*Hi*) described in Section 4.2.
2. Manually translate *Hi* into English (*En*). We refer to these English translations as *En(Manl.Trans.)*.
3. Automatically translate *Hi* into *En*. We refer to these English translations as *En(Auto.Trans.)*.
4. Manually annotate *Hi* for temporality. We call these *Hi(Manl.Tempo.)*.
5. Manually annotate all English datasets (*En(Manl.Trans.)* and *En(Auto.Trans.)*) for temporality. We call those *En(Manl.Trans., Manl.Tempo.)* and *En(Auto.Trans., Manl.Tempo)*, respectively.

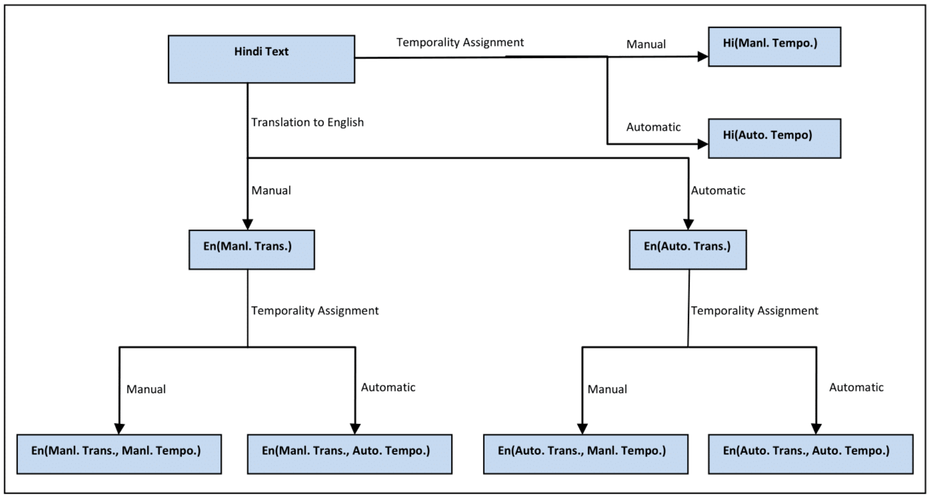

Figure 1: Proposed Architecture.

6. Run a Hindi temporality detector on *Hi*, creating *Hi(Auto.Tempo.)*

7. Run an English temporality detector on all the English datasets (*En(Manl.Trans.)* and *En(Auto.Trans.)*) creating *En(Manl.Trans., Auto.Tempo.)* and *En(Auto.Trans., Auto.Tempo.)*, respectively.

8. The procedural steps are depicted in Figure 1.

After creating various temporality-labeled datasets, we can compare the pairs of datasets to draw inferences. For example, comparison of the labels for *En(Manl.Trans., Manl.Tempo.)* and *En(Auto.Trans., Manl.Tempo.)* will show how the automatic translation affects the manual temporal levels with respect to the manual translation. The comparison will also show, for example, the extent to which a *past* sentence tends to be translated as a *present* sentence. The comparison of the dataset pairs (*Hi(Manl.Tempo.)* vs. *En(Auto.Trans., Auto.Tempo.)*) will show whether the idea of first translating a Hindi sentence into English and then using the automatic temporality detection is feasible or not. Section 5 demonstrates the procedure of Hindi to English translation. Section 6 describes the ways of finding temporality for different datasets i.e. *Hi*, *En(Manl.Trans.)* and *En(Auto.Trans.)*, both manually and automatically. Finally, Section 7 discusses the temporal error rate and analysis of different test cases.

4 Dataset

For our experiments, we use a parallel corpus of Hindi-English created in Bojar et al. (2014). This corpus contains 274k Hindi-English parallel sentences. The training and test sets for temporal tagging are described in Section 4.1 and 4.2. For MT, the details of training, test and development sets are mentioned in Section 5.

4.1 Training Set

We select past-, present-, and future-oriented texts using a manually selected high precision list of 50 seed terms.These are terms that capture temporal dimensions of texts with very few false positives, though the recall of these terms is low. In order to increase the recall, and to capture new terms that are good examples of past, present, and future, we expand our initial seed terms using a query expansion technique. For English, we use the publicly available word2vec[2] vectors that are trained on Google News corpus. For Hindi, we employ a continuously distributed vector representation of words using the continuous Skip-gram model (also known as Word2Vec) proposed by Mikolov et al. (2013) and trained on a corpus of around 44 million Hindi sentences developed by Bojar et al. (2014) with dimension set to 300 and window size set to 7.

Given the vector representations for the terms, we calculate the similarity scores between the pairs of terms in our vocabulary using cosine similarity. The top-10 similar terms for each seed term are selected for the expansion of the initial seed list. We again filter the whole collection of texts using the newly added seed terms. Table 1 shows few examples of expanded terms for some of the initial seed terms. There are some unrelated keywords in the expanded seed list due to the automatic process of keyword selection.

	Temporality	**Initial Seeds**	**Expanded Seed Terms**
Hindi	Past	गत (gata-past)[3]	विगत (vigata-last/past), पिछले (piChale-last/previous), बीते (bIte-past/bygone), पिछ्ले (piChalle-last/previous), विगत (vigata-last/past), गतवर्ष (gatavarSha-last year)
	Present	फ़िलहाल (phailahAla-at the moment)	फ़िलहाल (pḥailahAla-at the moment), अभी (abhI-now), अब (aba-now), फिलवक्त (philavakta-philanthropy), बहरहाल (baharahAla-nevertheless), ख़ैर (kḥaaira-well), हाल-फिलहाल (hAla-philahAla-most recently)
	Future	वादा (vAdA-promise)	वादे (vAde-promises), वायदा (vAyadA-futures), ऐलान (ailAna-announce), एलान (elAna-announce), दावा (dAvA-claim), आग्रह (Agraha-request)
English	Past	yesterday	yesterday, Earlier, Last, Shortly_afterwards, Meanwhile
	Present	currently	presently, Currently, now, currenty, still, already, iscurrently, actively
	Future	promise	promises, pledge, vow, commitment, hope, expect, vowing

Table 1: Examples of initial seed terms and their expanded terms.

Following this procedure, we create datasets for both Hindi and English containing 40K sentences each. Finally, we create our training set of 15k for both Hindi and English separately,[4] which consists of equally distributed past, present and future sentences. For the similar reason justified in Schwartz et al. (2015), we only considered past, present and future categories. Some example sentences are:

- नासा ने कल्पना के नाम से एक सुपर कंप्यूटर समर्पित किया है *(nAsA ne kalpanA ke nAma se eka supara kaMpyUTara samarpita kiyA hai-NASA has dedicated a super computer in the name of Kalpana)*, ***past***.

- अब ये अपने दोस्तों को बुलाने लगा है *(aba ye apane dostoM ko bulAne lagA hai-Now he is calling his friends)*, ***present***.

[2]https://code.google.com/p/word2vec/

[3]Henceforth, all the Hindi examples are represented by Hindi texts, ITRANS representations and using equivalent English translations.

[4]As our aim is to check whether temporality changes after translation or not, we are not using the translated version of the Hindi to create the training set for English.

- मेरे फूल को क्षण-भर में नष्ट हो जाने का जोखिम है *(mere phUla ko kShaNa-bhara meM naShTa ho jAne kA jokhima hai-My flower is at risk of being destroyed momentarily)*, ***future***.

4.2 Test Set

At first, we manually annotate the Hindi sentences with appropriate temporal categories from the same Hindi-English Bojar corpus. We made it sure that no training instances are being included. Finally, we select 996 sentences of past, present and future temporal classes. We call these 996 Hindi sentences *Hi* and the manually tagged *Hi* as *Hi(Manl. Tempo.)*. We then consider the manually translated English sentences(*En(Manl. Trans.)*) from *Hi* and then manually annotate them for temporality. We call these *En(Manl. Trans., Manl. Tempo.)*. We then manually annotate the automatically translated English sentences (*En(Auto. Trans.)*) from *Hi* for temporality. We call them as *En(Auto. Trans., Manl. Tempo.)*. Finally, we obtain three temporality tagged test sets, namely *Hi(Manl. Tempo.)*, *En(Manl. Trans., Manl. Tempo.)* and *En(Auto. Trans., Manl. Tempo.)*. We use *Hi* as the test set in Section 5 for MT.

5 Translation of Hindi to English

Our Hindi-English translation system, a phrase-based statistical MT system (Koehn et al., 2003), was built using Hindi-English parallel Bojar corpus (Bojar et al., 2014). We first remove the set (*Hi*) described in Section 4.2 from the corpus which is used as the test set for our MT system. We thereafter randomly select training and development sets from the rest of the corpus.

Set	#Sentences	#Tokens	
		En	Hi
Train	260,711	2,993,765	3,281,273
Test	996	23,806	27,012
Development	1000	12,480	14,153

Table 2: Statistics of data sets used in Hindi-English MT system

We tokenize, true-case and remove longer sentences as part of the preprocessing of the data. English sentences are tokenized using the *tokenizer.perl*[5] script, and we used the Indic_NLP_Library[6] for tokenizing Hindi sentences. After preprocessing, the training and development sets contain 260,711 and 1,000 parallel sentences, respectively. Details of the data sets are shown in Table 2.

For training, we use the Moses (Koehn et al., 2007) SMT system. We use KenLM (Heafield, 2011) for building a 4-gram language model and GIZA++ (Och and Ney, 2003) with the grow-diag-final-and heuristic for extracting phrases from the parallel corpus. The trained system is tuned using Minimum Error Rate Training (Och, 2003). For other parameters of Moses, default values are used. Automatic evaluation of our translation system achieves a BLEU (Papineni et al., 2002) score of 16.66.

6 Temporal Tagging of Sentences

We detect temporality in one Hindi dataset (*Hi*) and two English datasets *En(Manl. Trans.)*, *En(Auto. Trans.)* which denote manual and automatic translations from Hindi

[5] `https://github.com/moses-smt/mosesdecoder/blob/RELEASE-3.0/scripts/tokenizer/tokenizer.perl`

[6] `https://bitbucket.org/anoopk/indic\_nlp\_library`

to English language, respectively, as described in Section 4. We deploy both manual as well as automatic methods for temporal tagging.

6.1 Manual Temporal Tagging of Sentences

We create the datasets following manual annotation process as described in Section 4.2. Three annotators were asked to annotate based on the time sense in the sentences using past, present and future temporal categories. For the Hindi dataset(*Hi*), we considered only the temporal sentences, namely past, present and future. While annotating the two English datasets (*En(Manl. Trans.)*, *En(Auto. Trans.)*), we consider another category, namely *atemporal* apart from the three temporal categories. The reason for this consideration was to verify our hypothesis as to whether temporality is lost after translation. Finally, we consider sentences based on majority voting. We did not stick to the tense-based tagging as it sometimes misled the annotators to detect the actual temporality of the sentence. For example, consider the following sentence:

- आगामी छुट्टियों के लिए मेरे पास एक अच्छी योजना है *(AgAmI ChuTTiyoM ke lie mere pAsa eka achChI yojanA hai-I have a nice plan for the upcoming holidays).*

Here the tense of the verb *"have"* is *present* while the time sense of the sentence refers to *"future"*. Annotations also vary from person to person as any concrete definition of words does not exist; rather it is defined by the context appearing in the sentence. Finally, we obtain three sets of manually annotated datasets, namely *Hi(Manl. Tempo.)*, *En(Manl. Trans., Manl. Tempo.)* and *En(Auto. Trans., Manl. Tempo.)*. The temporality statistics are depicted in Table 3.

Datasets	Temporality(%)		
	Past	Present	Future
Hi(Manl. Tempo.)	32.83	24.80	42.37
En(Manl. Trans., Manl. Tempo.)	38.95	19.58	32.93
En(Auto. Trans., Manl. Tempo.)	41.15	11.75	34.74

Table 3: Class distribution of the manually annotated temporal datasets

From the statistics in Table 3, we can see that even after manual translation, loss of temporality is possible. The amount of loss in temporality in the dataset *En(Manl. Trans., Manl. Tempo.)* is 8.54%. In the automatically translated dataset *En(Auto. Trans., Manl. Tempo.)* the amount of loss in temporality is 13.35%, which is more than that of the manually translated set. Examples of these two cases are as follows:

1. **Manual Translation:** The temporality of the Hindi sentence "मानसिक रोग संबंधित लक्ष्य *(mAnasika roga saMbaMdhita lakShya)"* is *future*, but in the manually translated sentence *"Mental illness targets"*, the annotators tag it as *atemporal*. We observe that in the manually translated set, the temporality loss is mainly due to the incorrect temporal annotation rather than the incorrect manual translation. One of the possible reasons may be that the annotators were instructed not to see the temporal class of the Hindi sentence while labeling the English side. This was done to reduce bias.

2. **MT:** The Hindi sentence "ग्रामीण चीन में आर्थिक नवीनीकरण हुये हैं *(grAmINa chIna meM Arthika navInIkaraNa huye haiM- Economic Renewal happened in Rural China)"*, which has temporality *past*. This sentence is automatically translated as *"in rural areas are bound to China"* which becomes a factual text with no temporal sense. From our observation, we can say that the loss of temporality, in this

case, is mostly because of the wrong automatic translation rather than the the wrong manual annotation.

6.2 Automatic Temporal Classifier

We use a supervised machine learning-based approach for automatic sentence-level temporal tagging. For this experiment, we use the training set and test set as described in Section 4. We automatically classify three datasets, namely *Hi*, *En(Manl. Trans.)*, and *En(Auto. Trans.)*, for temporality in one of the three temporal categories, namely past, present or future. We employ one-vs.-rest approach for both our generation models as well as for evaluation. Our test set construction follows the same approach. For classification, we use Support Vector Machine (Joachims, 2002) classifier with word-unigram as a feature. Classification yields three sets of temporal datasets, named as *Hi(Auto. Tempo.)*, *En(Manl. Trans., Auto. Tempo.)* and *En(Auto. Trans., Auto. Tempo.)*. The class distribution of these temporal datasets is shown in Table 4.

Datasets	**Temporality(%)**		
	Past	**Present**	**Future**
Hi(Auto. Tempo.)	32.96	30.53	36.51
En(Manl. Trans., Auto. Tempo.)	16.12	20.97	62.91
En(Auto. Trans., Auto. Tempo.)	19.56	13.15	67.28

Table 4: Class distribution of the automatically tagged temporal datasets

7 Temporality after Translation

We generate all the manually and automatically labeled datasets mentioned in the experimental setup in Section 3 using the methods and systems described in Sections 3, 5 and 6. Results of class distribution in Table 3 can be compared with that in Table 4. The comparison of temporality labels between different data pairs is depicted in Table 5.

Data Pair	**Match(%)**
a. Hi(Manl. Tempo.) - Hi(Auto. Tempo.)	72.39
b. Hi(Manl. Tempo.) -En(Manl. Trans., Manl. Tempo.)	67.47
c. Hi(Manl. Tempo.) - En(Manl. Trans., Auto. Tempo.)	66.42
d. Hi(Manl. Tempo.) - En(Auto. Trans., Manl. Tempo.)	59.33
e. Hi(Manl. Tempo.) - En(Auto. Trans., Auto. Tempo.)	62.49
f. En(Manl. Trans., Manl. Tempo.) - En(Auto. Trans., Manl. Tempo.)	62.35
g. En(Manl. Trans., Manl. Tempo.) - En(Manl. Trans., Auto. Tempo.)	69.59
h. En(Auto. Trans., Manl. Tempo.) - En(Auto. Trans., Auto. Tempo.)	69.17

Table 5: Percentage of matching between pairs of temporality labeled datasets.

Row a., in Table 5 shows that the match percentage between the manual temporality and automatic temporality of Hindi texts is 72.39% which is the accuracy of the automatic temporality analysis system for Hindi.

Row b. shows the percentage match between the two manually temporal tagged datasets (*Hi(Manl.Tempo.)* and *En(Manl. Trans., Manl.Tempo.)*). We observe that two labels match only 67.47% of the time. It shows that the English translation does affect temporality.

Row c. shows the temporality match between the automatic temporality on manually translated texts and *Hi(Manl.Tempo.)*. Observe that the match for this pair is

66.42%, which is not too much lower than 67.47% obtained in the case of manual temporal tagging. This shows that English temporal system performs rather well. More importantly, the English automatic temporality analysis on the automatically translated texts shows a match of 62.49% (row e.), which makes this choice feasible for the temporality analysis of non-English texts.

Rows d. and e. show the temporality match of Hindi manual temporality with manual and automatic temporal labeling of the automatically translated texts, respectively. As the translation is automatic here, we expect these match percentages to be lower than those in rows b. and c. where the translation is manual, and the results show the same. However, we unexpectedly find the number for row e. to be higher than that of row d. This shows that some characteristics of the automatically translated text mislead humans with regards to the true temporality of the source text. However, this claim needs further insight in future.

Row f. shows the match between the manual temporal labels of manual and automatic translated English texts which is only 62.35%. Row g. shows the accuracy of the English automatic temporal analysis system when the translation is manual. The result of 69.59% shows that the quality of the English temporal analysis system is good, irrespective of human errors.

Row h. shows the accuracy of the English automatic temporal analysis system when the translation is automatic. In this case, the system's accuracy of 69.17% again shows that MT greatly impacts temporality.

We manually examine several Machine translated texts to understand the reason for incorrect annotations by humans with respect to Hindi annotation (row d. of Table 5). Most cases were due to translation errors where the temporal words were either lost or replaced by the other temporal words. Table 6 shows some examples of possible error cases. We observe that often the linking verb changes to a linking verb of a different temporality. In some cases, due to the change in the structure of the sentence, the temporality changes. Temporality loss happens mainly for the loss of action words and it occurs for all types of temporal sentences (past, present and future).

MT Error	**Temporality**
Change of linking verb after translation: **Hindi text:** तब लोगों को और अधिक बदला लेने की संभावना थी (taba logoM ko aura adhika badalA lene kI saMbhAvanA thI - Then people were more likely to take revenge.) **MT output:** when people are more likely to take revenge.	past future
Change of linking verb after translation: **Hindi text:** मैं अपनी नियति का पीछा कर रहा हूं (maiM apanI niyati kA pIChA kara rahA hUM - I'm following my destiny.) **MT output:** I was in pursuit of his destiny.	present past
Structural Change after translation: **Hindi text:** हमें उनकी प्रगति दिखाईये (hameM unakI pragati dikhAIye - Show us their progress.) **MT output:** their progress shows us.	future present
Loss of action word: **Hindi text:** लेकिन सचाई शायद कुछ और निकले (lekina sachAI shAyada kuCha aura nikale - But maybe a different truth can come out) **MT output:** but the truth and perhaps some.	future atemporal

Table 6: Examples of temporality change or loss due to different types of MT errors.

We analyze two cases to understand whether automatic temporality detection can be effective over the manual temporality in the translated instances. Our first case is comprised of the results in row b. and row c. of Table 5, where the translation is manual. There are some instances where the automatic temporality on the manually translated text correctly tags texts, while the manual temporality fails. The reason behind this is that the system can learn an appropriate model even from the mistranslated text. For example, consider the following case:

- **Hindi text:** "कि अगर किसी ने माना कि यह एक खतरा नहीं है *(ki agara kisI ne mAnA ki yaha eka khatarA nahIM hai)".*
- **Correct English translation:** *"That if somebody believes that this is not a threat".*
- **Manual English translation:** *"That if anybody believes that it wasn't such a threat".*

In the example, the temporal tag for the Hindi sentence is *future*, but when it is manually translated, the tag becomes *past*. In this case, the automatic English temporal tagger correctly predicts it as *future*. We observe that there are 6.7% instances in the manually translated English texts which are manually tagged incorrectly with respect to the Hindi text's temporality but correctly tagged by the automatic English temporal tagger.

Our second case is based on the results in row d. and row e. in Table 5, where the translation is automatic. The automatic temporal analysis system correctly tags several automatically translated instances (where manual labeling fails) for the same reason as for the first case. Consider the following examples:

- **Hindi text:** "तीसरी योजना में लगभग सभी अतिरिक्त क्षमता सार्वजनिक क्षेत्र को देते हुए इस्पात पिंडों का लक्ष्य 102 लाख टन पर निश्चित किया गया *(tIsarI yojanA meM lagabhaga sabhI atirikata kShamatA sArvajanika kShetra ko dete hue ispAta piMDoM kA lakShya 102 lAkha Tana para nishchita kiyA gayA)".*
- **Correct English translation:** *"Giving almost all the additional capacities to the public sector in the third plan, the goal of steel bodies was fixed at 102 lakh tonnes".*
- **MT output:** *"in the Third Plan the public sector almost all the additional capacity to steel ingots target of 102 million tonnes".*

In this case, the original temporal class in Hindi is *past*. In the machine translated English text, human experts annotate it as *future*, but the automatic temporal tagger tags it correctly as *past*. This case is quite interesting as despite obtaining some ungrammatical and unstructured sentences using machine translation, the automatic temporal tagger still correctly predicts temporality for some sentences. Our analysis shows that 8.14% temporal instances appear in the automatically translated English texts which are manually tagged incorrectly with respect to Hindi texts but correctly tagged by the automatic English temporal tagger.

8 Conclusion

In this paper, we present a case study on how machine translation affects temporality when the text is translated from Hindi to English. To the best of our knowledge this is the first study that systematically analyses various aspects of temporality preservation after translation. We create benchmark setups by creating manually labeled datasets for various test case scenarios. Our thorough investigation shows that temporality can be both lost and altered while text is translated from one source to the other target language. We also observe that the accuracy of the automatic temporal tagger in the automatically translated texts produces competitive results with respect to the accuracy of the automatic temporal tagger in the manually translated texts.

In future, we will explore these possible cases and further determine whether temporality preservation can improve the translation quality or not. We also propose to extend our study to more language pairs and use neural MT system for translation.

Acknowledgments

This research is partially supported by ADAPT Centre for Digital Content Technology, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

Adams, J. and Nettle, D. (2009). Time Perspective, Personality and Smoking, Body Mass, and Physical Activity: An Empirical Study. *British journal of health psychology*, 14(1):83–105.

Anand, A., Bedathur, S. J., Berberich, K., and Schenkel, R. (2012). Index Maintenance for Time-travel Text Search. In *The 35th International ACM SIGIR conference on research and development in Information Retrieval*, pages 235–244, Portland, OR, USA.

Ananthakrishnan, R., Kavitha, M., Jayprasad, J. H., Shekhar, R., and Bade, S. (2006). MaTra: A Practical Approach to Fully-automatic Indicative English-Hindi Machine Translation. In *Symposium on Modeling and Shallow Parsing of Indian Languages*, pages 1–8, Mumbai, India.

Bojar, O., Diatka, V., Rychlý, P., Stranák, P., Suchomel, V., Tamchyna, A., and Zeman, D. (2014). HindEnCorp - Hindi-English and Hindi-only Corpus for Machine Translation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 3550–3555, Reykjavik, Iceland.

Campos, R., Dias, G., Jorge, A. M., and Jatowt, A. (2014). Survey of Temporal Information Retrieval and Related Applications. *ACM Comput. Surv.*, 47(2):15:1–15:41.

Dave, S., Parikh, J., and Bhattacharyya, P. (2001). Interlingua-based English–Hindi Machine Translation and Language Divergence. *Machine Translation*, 16(4):251–304.

Dias, G. H., Hasanuzzaman, M., Ferrari, S., and Mathet, Y. (2014). TempoWordNet for Sentence Time Tagging. In *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, pages 833–838, Geneva, Switzerland.

Dungarwal, P., Chatterjee, R., Mishra, A., Kunchukuttan, A., Shah, R., and Bhattacharyya, P. (2014). The IIT Bombay Hindi-English Translation System at WMT 2014. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 90–96, Baltimore, Maryland, USA.

Heafield, K. (2011). KenLM: Faster and Smaller Language Model Queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 187–197, Edinburgh, Scotland,.

Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms.* Kluwer Academic Publisher, Dordrecht, Netherlands.

Joho, H., Jatowt, A., Blanco, R., Naka, H., and Yamamoto, S. (2014). Overview of NTCIR-11 Temporal Information Access (Temporalia) Task. In *Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies*, pages 429–437, Tokyo, Japan.

Kanhabua, N., Blanco, R., and Matthews, M. (2011). Ranking Related News Predictions. In *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 755–764, Beijing, China.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., et al. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, pages 177–180, Prague, Czech Republic.

Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical Phrase-based Translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54, Edmonton, Canada.

Kulkarni, A., Teevan, J., Svore, K. M., and Dumais, S. T. (2011). Understanding Temporal Query Dynamics. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining*, pages 167–176, Hong Kong, China.

Lohar, P., Afli, H., and Way, A. (2017). Maintaining Sentiment Polarity in Translation of User-Generated Content. *The Prague Bulletin of Mathematical Linguistics*, 108(1):73–84.

Mani, I., Pustejovsky, J., and Gaizauskas, R. (2005). *The Language of Time: a Reader*, volume 126. Oxford University Press, Oxford, UK.

Metzger, M. (2007). Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research. *Journal of the American Society for Information Science and Technology*, 58(13):2078–2091.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Advances in neural information processing systems*, pages 3111–3119, Lake Tahoe, Nevada, United States.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, 38(11):39–41.

Och, F. J. (2003). Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 160–167, Sapporo, Japan.

Och, F. J. and Ney, H. (2003). A Systematic Comparison of Carious Statistical Alignment Models. *Computational linguistics*, 29(1):19–51.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318, Philadelphia, Pennsylvania.

Pawar, D., Hasanuzzaman, M., and Ekbal, A. (2016). Building Tempo-HindiWordNet: A Resource for Effective Temporal Information Access in Hindi. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016*, pages 3752–3759, Portorož, Slovenia.

Ramanathan, A., Hegde, J., Shah, R. M., Bhattacharyya, P., and Sasikumar, M. (2008). Simple Syntactic and Morphological Processing Can Help English-Hindi Statistical Machine Translation. In *Third International Joint Conference on Natural Language Processing*, pages 513–520, Hyderabad, India.

Sachdeva, K., Srivastava, R., Jain, S., and Sharma, D. M. (2014). Hindi to English Machine Translation: Using Effective Selection in Multi-Model SMT. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 1807–1811, Reykjavik, Iceland.

Salameh, M., Mohammad, S., and Kiritchenko, S. (2015). Sentiment after Translation: A Case-Study on Arabic Social Media Posts. In *Proceedings of The 2015 Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies (NAACL)*, pages 767–777, Denver, Colorado, USA.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., et al. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PloS one*, 8(9):1–16.

Schwartz, H. A., Park, G., Sap, M., Weingarten, E., Eichstaedt, J., Kern, M., Berger, J., Seligman, M., and Ungar, L. (2015). Extracting Human Temporal Orientation in Facebook Language. In *Proceedings of The 2015 Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies (NAACL)*, pages 409–419, Denver, Colorado, USA.

Sen, S., Banik, D., Ekbal, A., and Bhattacharyya, P. (2016). IITP English-Hindi Machine Translation System at WAT 2016. In *Proceedings of the 3rd Workshop on Asian Translation (WAT)*, pages 216–222, Osaka, Japan.

Sinha, R. and Jain, A. (2003). AnglaHindi: an English to Hindi Machine-aided Translation System. In *MT Summit IX*, pages 494–497, New Orleans, USA.

Sinha, R. and Thakur, A. (2005). Machine Translation of Bi-lingual Hindi-English (Hinglish) Text. In *10th Machine Translation summit (MT Summit X)*, pages 149–156, Phuket, Thailand.

UzZaman, N., Llorens, H., Derczynski, L., Allen, J. F., Verhagen, M., and Pustejovsky, J. (2013). SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. In *Proceedings of the 7th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2013*, pages 1–9, Atlanta, Georgia, USA.

Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Moszkowicz, J., and Pustejovsky, J. (2009). The TempEval Challenge: Identifying Temporal Relations in Text. *Language Resources and Evaluation*, 43(2):161–179.

Verhagen, M., Saurí, R., Caselli, T., and Pustejovsky, J. (2010). SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval@ACL 2010*, pages 57–62, Uppsala, Sweden.

Webley, P. and Nyhus, E. K. (2006). Parents' Influence on Children's Future Orientation and Saving. *Journal of Economic Psychology*, 27(1):140–164.

Zimbardo, P. G. and Boyd, J. N. (2015). Putting Time in Perspective: A Valid, Reliable Individual-Differences Metric. In *Time Perspective Theory; Review, Research and Application*, pages 17–55. Springer, Berlin, Germany.